

## **Crowdsourced collection and analysis of spoken data**

*Rupal Patel, Michael Suen, Geoffrey Meltzner*

Studies of speech motor control and motor speech disorders require access to spoken corpora. Historically, these data have been gathered in highly controlled laboratory settings using specialized equipment. The proliferation of high fidelity recording equipment within mobile devices and the reduced cost of data storage provides an unprecedented opportunity to transform the study of speech production and its impairment. The Human Voicebank initiative is a global, crowdsourced effort with over 25,000 speakers ranging in age from 6-91, from 110 countries who have collectively contributed 9.7M+ recordings. The web platform was initially designed to gather spoken corpora for use in the development of personalized digital voices. This talk will describe the advantages and challenges of the corpus and plans to make a portion of it available to the scientific community. We hope it will also spark a conversation about collaborations and extensions to the platform to fuel future innovation.